



Linked Open Data to Qualify, Enrich and Validate Experiments: The Semantic Approach to Integrated Data is Bearing Fruits

Erich Gombocz^{1)*}, Robert Stanley¹⁾

¹⁾IO Informatics, Inc., Berkeley, CA, USA.

*Correspondence: egombocz@io-informatics.com

Abstract:

Semantic methods fulfill dynamic integration requirements and provide the framework for rich semantic queries (SPARQL) to answer complex biological questions. While it has been known for years, that semantic methods make it possible for domain experts, ontologists and informaticians, to quickly build, modify and extend integrated knowledge bases, such methods have become broadly invaluable today. The fact, that semantic W3C standards like RDF assure coherence, harmonized synonyms and terminologies, and provide an extensible data integration framework and interactive knowledgebase for reasoning and relevant network analysis has driven their increased use in industry and academia..

A use case of toxicology studies performed on different experimental platforms (gene expression and metabolic profiling) is demonstrated. Building integrated knowledge - harmonized and enriched with data from Drugbank, Disasome, SIDER, UniProt, NCBI Biosystems – helps with toxic assessment in drug development, classification of toxicity types (hepatotoxicity, nephrotoxicity, categorical toxicity based on drug residues) and provides better a-priori determination of adverse effects of drug combinations. With the ever growing number of quality resources joining the Linked Open Data cloud, more data become publicly available to extend understanding of experimental data through enrichment with their biological systems-induced mechanism in ways never anticipated before. Their effective use exemplifies the success of using an innovative semantic approach to integrate all experimental, internal, external, clinical and public data sources. The resulting visual exploration of such an integrated graph environment and the construction of characteristic marker patterns or molecular signatures are applicable to predictive biology-based decision support for complex translational research and personalized medicine applications. Now drug targets, interactions, adverse effects can contextualize experiments through public resource-enriched biological qualification and validation. Standards-based semantic integration pays off in a big way and paves the way to the development of next generation drugs with understood toxicity, efficacy and minimal side effects.

Audience:

Pharma research, translational research, clinical research facilities

On a real-life example, the use of RDF-based linked open data in drug discovery and translational research is exemplified. It will be shown how public information about drug targets, interactions and adverse effects information can contextualize experiments to qualify results and

to pave the way to the development of next generation drugs with understood toxicity, efficacy and minimal side effects.

Biography:

Dr. Erich Gombocz has over 30 years of experience in Life Science research, laboratory automation and data management in scientific and distributed systems environments. His over 30 years programming skills in instrument control, user interface, database design, scientific analysis, and on-line laboratory automation and as developer of innovative software algorithms and architecture, he co-founded IO Informatics in 2003 with focus on semantic data integration and knowledge management in life sciences to apply systems biology approaches to pharmaceutical and clinical decision-making. Dr. Gombocz has published over 60 scientific publications and holds more than 40 biotechnology- and software-related US and international patents. Considered an international expert in separation science and bioinformatics, he serves on the editorial board of several scientific journals. His activities in several World Wide Web Consortium (W3C) HCLS Special Interest Groups, at the National Center for Biomedical Ontology (NCBO), the Pistoia Alliance Standards Initiative and as Chair of the Working Group for Best Practices in Data Sharing are a testimonial for his role at the forefront of technology.
